

# Weekly Report

Lu Junhua

2015 年 10 月 11 日

These days, I devoted most of my time in processing data of netease players who buy shizhuang. I write a document for all the datatable in my local database. They are attached in the end of this file. While processing the data, I learned many tricks:

- create index on tables will profoundly increase the efficiency
- `os.listdir` in python can list all the files under a contents and put them in a list
- in mysql, load data has many parameters like “fields terminated by [optionally] enclosed by escaped by lines terminated by”, etc
- How to find all the repeated item in one table? using group by all the attribute. like:

```
select * from netease.buy_shizhuang group by date_time , serverid ,  
CostumeID , buyerid , cost , Quantity , userid , isgift  
having count(date_time and serverid and CostumeID and buyerid  
and Quantity and userid and isgift)>1
```

- Again hashtable is efficient.

And there also several findings in data self.

- Every time one ID logs in, it will be recorded. One id can create up to 4 roles. Their gender can be different. However, it will cost lots of money and only a few change their gender.

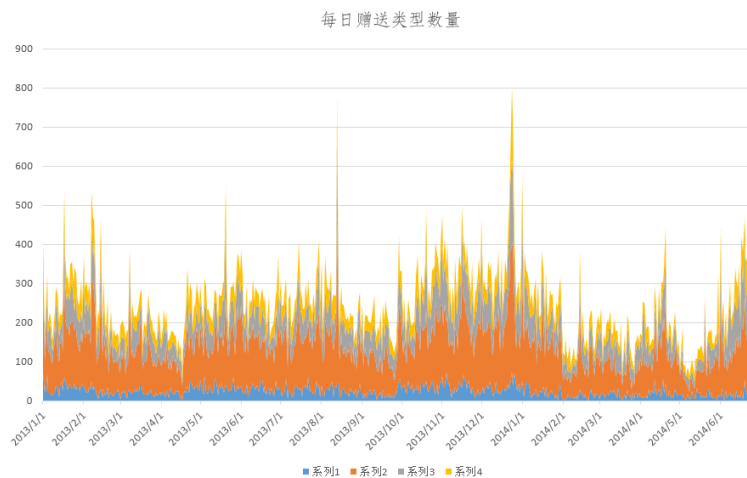
## 角色创建中一些统计

创建次数	4	3	2	1
人次	1969	24031	16822	176529
有换性别的人次	15	54	33	/
换占比例	0.76%	0.22%	0.2%	/

注意到那些四次的几乎都在2014.6.26日有创建过两个角色. 那天是有个合服记录. 合服记录可以在官方论坛:  
<http://xqn.netease.com/forum-11-1.html> 开发日志区或者  
<http://xqn.netease.com/forum.php?mod=viewthread&tid=975542&highlight=合服>

- If someone log on day1, and not log off. And he has NO log-on records on day1+1,day1+2... until log off and log on again.

## 不同性别类型的赠送分布



And there are problems with data.

- 5 repeated rows in buy\_shizhuang. Why?
- Logon records missing 2013.1.8–2013.1.31 and 2013.2.25–2013.3.5

The following are things to be done.

- Data Extraction for attributes before 1st buy shizhuang using parallel coordinates.
- Network extraction using python networkx package. I'm thinking about how to do it by myself and today got to know that networkx can make it easy.

## 时装购买者的几张属性表的说明

陆俊华

一些简要说明:

- 以下的“所有 id”都是在 所有 buy\_shizhuang 中出现的 id(不管是购买者还是使用者)
- buy\_shizhuang 中记录的时间跨度为 2013.1.1~2014.6.30
- 在我本机存储的都在 MySQL 数据库中

下面是详细情况:

<ftp://www.cad.zju.edu.cn>    账户: netease    密码: cadnetcgease

外网应该需要 rvpn 访问

/buy\_shizhuang/create\_gender.cs

大小	条数	时间跨度
11.2M	290,142	2013.1.1~2014.6.30
内容是: 所有 id 的创建角色性别 玩家 id, 服务器 id, 时间(年月日时分秒), 角色性别 一个 id 可以创建多个角色, 所以可能是不同的性别 另外据悉可以花 60 人民币改变角色性别		

在数据库中的存储是:

playerid	varchar(15)
serverid	smallint(4)
date_time	datetime
gender	tinyint(4)

/buy\_shizhuang/s2y\_selected.csv

大小	条数	时间跨度
549M	7,173,759	2013.3.6~2014.6.30 缺 2014-01-03 应该是遗漏了
内容是: 所有 id 的银两换元宝记录 记录编号(主键), 时间, 服务器 id, 银两, 元宝, 银两->元宝兑换率, 卖家 id, 买家 id		

在数据库中存储是:

recid	Int(11)
date_time	datetime

serverid	smallint(4)
silver	int(11)
yuanbao	int(11)
s2y_rate	double
sellerid	varchar(15)
buyerid	varchar(15)

/buy\_shizhuang/y2s\_selected.csv

大小	条数	时间跨度
2.91G	42,620,867	2013.3.6~2014.6.30 齐
内容是: 所有 id 的银两换元宝记录 记录编号(主键), 时间, 服务器 id, 元宝, 银两, 元白->银两兑换率, 卖家 id, 买家 id		

在数据库中存储是:

recid	Int(11)
date_time	datetime
serverid	smallint(4)
yuanbao	int(11)
silver	int(11)
y2s_rate	double
sellerid	varchar(15)
buyerid	varchar(15)

/buy\_shizhuang /logon\_selected.csv

大小	条数	时间跨度
8.23G	161,431,575	2013.1.1~2014.6.30
内容是: 所有 id 的登陆记录, 一天可以登陆多次, 每次都会留下记录 时间, 服务器 id, 玩家 id, 游戏中等级, IP 地址		

在数据库中存储是:

date_time	datetime
serverid	smallint(4)
playerid	varchar(15)
game_level	smallint(4)
ip_addr	varchar(16)

/buy\_shizhuang/player\_daily\_attr\_selected.csv

大小	条数	时间跨度
1.02G	22,109,727	2013.5.24~2014.6.30
内容是: 所有 id 每天当天首次上线时间属性记录 玩家 id, 服务器 id, 时间, 装备等级, 战斗力, 修为, 修炼 2013.8.15 以前都没有战斗力这个指标 2013.5.17~2013.5.23 文件都是空文件		

在数据库中存储是

playerid	varchar(15)
serverid	smallint(4)
date_time	datetime
zhuangbei_level	smallint(2)
zhandouli	varchar(10) 因为出现了 NULL
xiuwei	smallint(6)
xiulian	smallint(6)

/buy\_shizhuang/关于聊天的记录(并没有聊天的内容). 还在传, 很大, 大概么 150G 样子.

大小	条数	时间跨度
约 120G	2,549,472,400	2013.1.1~2014.6.30
内容是: 时间, 服务器 id, 发送者 id, 接收者 id. 每次聊天都会留下记录, 并且没有聊天内容		

在数据库中存储是:

date_time	datetime
serverid	smallint(4)
sendid	varchar(15)
receiveid	varchar(15)